

Sparse approximation by greedy algorithms

V. Temlyakov *

Abstract

It is a survey on recent results in constructive sparse approximation. Three directions are discussed here: (1) Lebesgue-type inequalities for greedy algorithms with respect to a special class of dictionaries, (2) constructive sparse approximation with respect to the trigonometric system, (3) sparse approximation with respect to dictionaries with tensor product structure. In all three cases constructive ways are provided for sparse approximation. The technique used is based on fundamental results from the theory of greedy approximation. In particular, results in the direction (1) are based on deep methods developed recently in compressed sensing. We present some of these results with detailed proofs.

1 Introduction

The paper is a survey on recent breakthrough results in constructive sparse approximation. In all cases discussed here the new technique is based on greedy approximation. The main motivation for the study of sparse approximation is that many real world signals can be well approximated by sparse ones. Sparse approximation automatically implies a need for nonlinear approximation, in particular, for greedy approximation. We give a brief description of a sparse approximation problem and present a discussion of the obtained results and their relation to previous work. In Section 2 we concentrate on breakthrough results from [18] and [41]. In these papers we

*University of South Carolina and Steklov Institute of Mathematics. Research was supported by NSF grant DMS-1160841

extended a fundamental result of Tong Zhang [48] on the Lebesgue-type inequality for the RIP dictionaries in a Hilbert space (see Theorem 2.2 below) in several directions. We found new more general than the RIP conditions on a dictionary, which still guarantee the Lebesgue-type inequalities in a Hilbert space setting. We generalized these conditions to a Banach space setting and proved the Lebesgue-type inequalities for dictionaries satisfying those conditions. To illustrate the power of new conditions we applied this new technique to bases instead of redundant dictionaries. In particular, this technique gave very strong results for the trigonometric system.

In a general setting we are working in a Banach space X with a redundant system of elements \mathcal{D} (dictionary \mathcal{D}). There is a solid justification of importance of a Banach space setting in numerical analysis in general and in sparse approximation in particular (see, for instance, [40], Preface, and [29]). An element (function, signal) $f \in X$ is said to be m -sparse with respect to \mathcal{D} if it has a representation $f = \sum_{i=1}^m x_i g_i$, $g_i \in \mathcal{D}$, $i = 1, \dots, m$. The set of all m -sparse elements is denoted by $\Sigma_m(\mathcal{D})$. For a given element f_0 we introduce the error of best m -term approximation $\sigma_m(f_0, \mathcal{D}) := \inf_{f \in \Sigma_m(\mathcal{D})} \|f_0 - f\|$. We are interested in the following fundamental problem of sparse approximation.

Problem. How to design a practical algorithm that builds sparse approximations comparable to best m -term approximations?

In a general setting we study an algorithm (approximation method) $\mathcal{A} = \{A_m(\cdot, \mathcal{D})\}_{m=1}^\infty$ with respect to a given dictionary \mathcal{D} . The sequence of mappings $A_m(\cdot, \mathcal{D})$ defined on X satisfies the condition: for any $f \in X$, $A_m(f, \mathcal{D}) \in \Sigma_m(\mathcal{D})$. In other words, A_m provides an m -term approximant with respect to \mathcal{D} . It is clear that for any $f \in X$ and any m we have $\|f - A_m(f, \mathcal{D})\| \geq \sigma_m(f, \mathcal{D})$. We are interested in such pairs $(\mathcal{D}, \mathcal{A})$ for which the algorithm \mathcal{A} provides approximation close to best m -term approximation. We introduce the corresponding definitions.

Definition 1.1. We say that \mathcal{D} is an almost greedy dictionary with respect to \mathcal{A} if there exist two constants C_1 and C_2 such that for any $f \in X$ we have

$$\|f - A_{C_1 m}(f, \mathcal{D})\| \leq C_2 \sigma_m(f, \mathcal{D}). \quad (1.1)$$

If \mathcal{D} is an almost greedy dictionary with respect to \mathcal{A} then \mathcal{A} provides almost ideal sparse approximation. It provides $C_1 m$ -term approximant as good (up to a constant C_2) as ideal m -term approximant for every $f \in X$. In the case $C_1 = 1$ we call \mathcal{D} a greedy dictionary. We also need a more general definition. Let $\phi(u)$ be a function such that $\phi(u) \geq 1$.

Definition 1.2. We say that \mathcal{D} is a ϕ -greedy dictionary with respect to \mathcal{A} if there exists a constant C_3 such that for any $f \in X$ we have

$$\|f - A_{\phi(m)m}(f, \mathcal{D})\| \leq C_3 \sigma_m(f, \mathcal{D}). \quad (1.2)$$

If $\mathcal{D} = \Psi$ is a basis then in the above definitions we replace dictionary by basis. Inequalities of the form (1.1) and (1.2) are called the Lebesgue-type inequalities.

In the above setting the quality criterion of the algorithm \mathcal{A} is based on the Lebesgue-type inequalities, which hold for every individual $f \in X$. In classical approximation theory very often we use as a quality criterion of the algorithm \mathcal{A} its performance on a given class F . In this case we compare

$$e_m(F, \mathcal{A}, \mathcal{D}) := \sup_{f \in F} \|f - A_m(f, \mathcal{D})\|$$

with

$$\sigma_m(F, \mathcal{D}) := \sup_{f \in F} \sigma_m(f, \mathcal{D}).$$

We discuss this setting in Sections 5 and 6.

In the case $\mathcal{A} = \{G_m(\cdot, \Psi)\}_{m=1}^\infty$ is the Thresholding Greedy Algorithm (TGA) the theory of greedy and almost greedy bases is well developed (see [40]). We remind that in case of a normalized basis $\Psi = \{\psi_k\}_{k=1}^\infty$ of a Banach space X the TGA at the m th iteration gives an approximant

$$G_m(f, \Psi) := \sum_{j=1}^m c_{k_j} \psi_{k_j}, \quad f = \sum_{k=1}^\infty c_k \psi_k, \quad |c_{k_1}| \geq |c_{k_2}| \geq \dots$$

In particular, it is known (see [40], p. 17) that the univariate Haar basis is a greedy basis with respect to TGA for all L_p , $1 < p < \infty$. Also, it is known that the TGA does not work well with respect to the trigonometric system.

We demonstrated in the paper [41] that the Weak Chebyshev Greedy Algorithm (WCGA) which we define momentarily is a solution to the above problem for a special class of dictionaries.

Let X be a real Banach space with norm $\|\cdot\| := \|\cdot\|_X$. We say that a set of elements (functions) \mathcal{D} from X is a dictionary if each $g \in \mathcal{D}$ has norm one ($\|g\| = 1$), and the closure of $\text{span } \mathcal{D}$ is X . For a nonzero element $g \in X$ we let F_g denote a norming (peak) functional for g : $\|F_g\|_{X^*} = 1$, $F_g(g) = \|g\|_X$. The existence of such a functional is guaranteed by the Hahn-Banach theorem.

Let $t \in (0, 1]$ be a given weakness parameter. We define the Weak Chebyshev Greedy Algorithm (WCGA) (see [37]) as a generalization for Banach spaces of the Weak Orthogonal Matching Pursuit (WOMP). In a Hilbert space the WCGA coincides with the WOMP. The WOMP is very popular in signal processing, in particular, in compressed sensing. In case $t = 1$, WOMP is called Orthogonal Matching Pursuit (OMP).

Weak Chebyshev Greedy Algorithm (WCGA). Let f_0 be given. Then for each $m \geq 1$ we have the following inductive definition.

- (1) $\varphi_m := \varphi_m^{c,t} \in \mathcal{D}$ is any element satisfying

$$|F_{f_{m-1}}(\varphi_m)| \geq t \sup_{g \in \mathcal{D}} |F_{f_{m-1}}(g)|.$$

- (2) Define $\Phi_m := \Phi_m^t := \text{span}\{\varphi_j\}_{j=1}^m$, and define $G_m := G_m^{c,t}$ to be the best approximant to f_0 from Φ_m .

- (3) Let $f_m := f_m^{c,t} := f_0 - G_m$.

The trigonometric system is a classical system that is known to be difficult to study. In [41] we study among other problems the problem of nonlinear sparse approximation with respect to it. Let \mathcal{RT} denote the real trigonometric system $1, \sin 2\pi x, \cos 2\pi x, \dots$ on $[0, 1]$ and let \mathcal{RT}_p to be its version normalized in $L_p([0, 1])$. Denote $\mathcal{RT}_p^d := \mathcal{RT}_p \times \dots \times \mathcal{RT}_p$ the d -variate trigonometric system. We need to consider the real trigonometric system because the algorithm WCGA is well studied for the real Banach space. In order to illustrate performance of the WCGA we discuss in this section the above mentioned problem for the trigonometric system. We proved in [41] the following Lebesgue-type inequality for the WCGA.

Theorem 1.1. *Let \mathcal{D} be the normalized in L_p , $2 \leq p < \infty$, real d -variate trigonometric system. Then for any $f_0 \in L_p$ the WCGA with weakness parameter t gives*

$$\|f_{C(t,p,d)m \ln(m+1)}\|_p \leq C \sigma_m(f_0, \mathcal{D})_p. \quad (1.3)$$

The Open Problem 7.1 (p. 91) from [38] asks if (1.3) holds without an extra $\ln(m+1)$ factor. Theorem 1.1 is the first result on the Lebesgue-type inequalities for the WCGA with respect to the trigonometric system. It provides a progress in solving the above mentioned open problem, but the problem is still open.

We note that properties of a given basis with respect to TGA and WCGA could be very different. For instance, the class of quasi-greedy bases (with

respect to TGA), that is the class of bases Ψ for which $G_m(f, \Psi)$ converges for each $f \in X$, is a rather narrow subset of all bases. It is close in a certain sense to the set of unconditional bases. The situation is absolutely different for the WCGA. If X is uniformly smooth then WCGA converges for each $f \in X$ with respect to any dictionary in X (see [40], Ch.6).

Theorem 1.1 shows that the WCGA is very well designed for the trigonometric system. We show in [41] that an analog of (1.3) holds for uniformly bounded orthogonal systems. The proof of Theorem 1.1 uses technique developed in compressed sensing for proving the Lebesgue-type inequalities for redundant dictionaries with special properties. First results on Lebesgue-type inequalities were proved for incoherent dictionaries (see [40] for a detailed discussion). Then a number of results were proved for dictionaries satisfying the Restricted Isometry Property (RIP) assumption. The incoherence assumption on a dictionary is stronger than the RIP assumption. The corresponding Lebesgue-type inequalities for the Orthogonal Matching Pursuit (OMP) under RIP assumption were not known for a while. As a result new greedy-type algorithms were introduced and exact recovery of sparse signals and the Lebesgue-type inequalities were proved for these algorithms: the Regularized Orthogonal Matching Pursuit (see [22]), Compressive Sampling Matching Pursuit (CoSaMP) (see [21]), and the Subspace Pursuit (SP) (see [3]). The OMP is simpler than CoSaMP and SP, however, at the time of invention of CoSaMP and SP these algorithms provided exact recovery of sparse signals and the Lebesgue-type inequalities for dictionaries satisfying the Restricted Isometry Property (RIP) (see [21] and [3]). The corresponding results for the OMP were not known at that time. Later, a breakthrough result in this direction was obtained by Zhang [48]. In particular, he proved that if \mathcal{D} satisfies RIP then the OMP recovers exactly all m -sparse signals within Cm iterations. In [18] and [41] we developed Zhang's technique to obtain recovery results and the Lebesgue-type inequalities in the Banach space setting.

The above Theorem 1.1 guarantees that the WCGA works very well for each individual function f . It is a constructive method, which provides after $\asymp m \ln m$ iterations an error comparable to $\sigma_m(f, \mathcal{D})$. Here are two important points. First, in order to guarantee a rate of decay of errors $\|f_n\|$ of the WCGA we would like to know how smoothness assumptions on f_0 affect the rate of decay of $\sigma_m(f_0, \mathcal{D})$. Second, if, as we believe, one cannot get rid of $\ln m$ in Theorem 1.1 then it would be nice to find a constructive method, which provides on a certain smoothness class the order of best m -term ap-

proximation after m iterations. Thus, as a complement to Theorem 1.1 we would like to obtain results, which relate rate of decay of $\sigma_m(f, \mathcal{T}^d)_p$ to some smoothness type properties of f . In Section 5 we concentrate on constructive methods of m -term approximation. We measure smoothness in terms of mixed derivative and mixed difference. We note that the function classes with bounded mixed derivative are not only interesting and challenging object for approximation theory but they are important in numerical computations.

We discuss here the problem of sparse approximation. This problem is closely connected with the problem of recovery of sparse functions (signals). In the sparse recovery problem we assume that an unknown function f is sparse with respect to a given dictionary and we want to recover it. This problem was a starting point for the compressed sensing theory (see [40], Ch.5). In particular, the celebrated contribution of the work of Candes, Tao and Donoho was to show that the recovery can be done by the ℓ_1 minimization algorithm. We stress that ℓ_1 minimization algorithm works for the exact recovery of sparse signals. It does not provide sparse approximation. The greedy-type algorithms discussed in this paper provide sparse approximation, satisfying the Lebesgue-type inequalities. It is clear that the Lebesgue-type inequalities (1.1) and (1.2) guarantee exact recovery of sparse signals.

2 Lebesgue-type inequalities. General results.

A very important advantage of the WCGA is its convergence and rate of convergence properties. The WCGA is well defined for all m . Moreover, it is known (see [37] and [40]) that the WCGA with weakness parameter $t \in (0, 1]$ converges for all f_0 in all uniformly smooth Banach spaces with respect to any dictionary. That is, when X is a real Banach space and the modulus of smoothness of X is defined as follows

$$\rho(u) := \frac{1}{2} \sup_{x, y; \|x\|=\|y\|=1} |\|x + uy\| + \|x - uy\| - 2|, \quad (2.1)$$

then the uniformly smooth Banach space is the one with $\rho(u)/u \rightarrow 0$ when $u \rightarrow 0$.

We discuss here the Lebesgue-type inequalities for the WCGA with weakness parameter $t \in (0, 1]$. This discussion is based on papers [18] and [41]. For notational convenience we consider here a countable dictionary $\mathcal{D} = \{g_i\}_{i=1}^\infty$. The following assumptions **A1** and **A2** were used in [18]. For a given f_0 let

sparse element (signal)

$$f := f^\epsilon = \sum_{i \in T} x_i g_i, \quad g_i \in \mathcal{D},$$

be such that $\|f_0 - f^\epsilon\| \leq \epsilon$ and $|T| = K$. For $A \subset T$ denote

$$f_A := f_A^\epsilon := \sum_{i \in A} x_i g_i.$$

A1. We say that $f = \sum_{i \in T} x_i g_i$ satisfies the Nikol'skii-type $\ell_1 X$ inequality with parameter r if

$$\sum_{i \in A} |x_i| \leq C_1 |A|^r \|f_A\|, \quad A \subset T. \quad (2.2)$$

We say that a dictionary \mathcal{D} has the Nikol'skii-type $\ell_1 X$ property with parameters K, r if any K -sparse element satisfies the Nikol'skii-type $\ell_1 X$ inequality with parameter r .

A2. We say that $f = \sum_{i \in T} x_i g_i$ has incoherence property with parameters D and U if for any $A \subset T$ and any Λ such that $A \cap \Lambda = \emptyset$, $|A| + |\Lambda| \leq D$ we have for any $\{c_i\}$

$$\|f_A - \sum_{i \in \Lambda} c_i g_i\| \geq U^{-1} \|f_A\|. \quad (2.3)$$

We say that a dictionary \mathcal{D} is (K, D) -unconditional with a constant U if for any $f = \sum_{i \in T} x_i g_i$ with $|T| \leq K$ inequality (2.3) holds.

The term *unconditional* in **A2** is justified by the following remark. The above definition of (K, D) -unconditional dictionary is equivalent to the following definition. Let \mathcal{D} be such that any subsystem of D distinct elements e_1, \dots, e_D from \mathcal{D} is linearly independent and for any $A \subset [1, D]$ with $|A| \leq K$ and any coefficients $\{c_i\}$ we have

$$\left\| \sum_{i \in A} c_i e_i \right\| \leq U \left\| \sum_{i=1}^D c_i e_i \right\|.$$

It is convenient for us to use the following assumption **A3** introduced in [41] which is a corollary of assumptions **A1** and **A2**.

A3. We say that $f = \sum_{i \in T} x_i g_i$ has ℓ_1 incoherence property with parameters D , V , and r if for any $A \subset T$ and any Λ such that $A \cap \Lambda = \emptyset$, $|A| + |\Lambda| \leq D$ we have for any $\{c_i\}$

$$\sum_{i \in A} |x_i| \leq V|A|^r \|f_A - \sum_{i \in \Lambda} c_i g_i\|. \quad (2.4)$$

A dictionary \mathcal{D} has ℓ_1 incoherence property with parameters K , D , V , and r if for any $A \subset B$, $|A| \leq K$, $|B| \leq D$ we have for any $\{c_i\}_{i \in B}$

$$\sum_{i \in A} |c_i| \leq V|A|^r \left\| \sum_{i \in B} c_i g_i \right\|.$$

It is clear that **A1** and **A2** imply **A3** with $V = C_1 U$. Also, **A3** implies **A1** with $C_1 = V$ and **A2** with $U = V K^r$. Obviously, we can restrict ourselves to $r \leq 1$.

We give a simple remark that widens the collection of dictionaries satisfying the above properties **A1**, **A2**, and **A3**.

Definition 2.1. Let $\mathcal{D}^1 = \{g_i^1\}$ and $\mathcal{D}^2 = \{g_i^2\}$ be countable dictionaries. We say that \mathcal{D}^2 D -dominates \mathcal{D}^1 (with a constant B) if for any set Λ , $|\Lambda| \leq D$, of indices and any coefficients $\{c_i\}$ we have

$$\left\| \sum_{i \in \Lambda} c_i g_i^1 \right\| \leq B \left\| \sum_{i \in \Lambda} c_i g_i^2 \right\|.$$

In such a case we write $\mathcal{D}^1 \prec \mathcal{D}^2$ or more specifically $\mathcal{D}^1 \leq B \mathcal{D}^2$.

In the case $\mathcal{D}^1 \leq E_1^{-1} \mathcal{D}^2$ and $\mathcal{D}^2 \leq E_2 \mathcal{D}^1$ we say that \mathcal{D}^1 and \mathcal{D}^2 are D -equivalent (with constants E_1 and E_2) and write $\mathcal{D}^1 \approx \mathcal{D}^2$ or more specifically $E_1 \mathcal{D}^1 \leq \mathcal{D}^2 \leq E_2 \mathcal{D}^1$.

Proposition 2.1. Assume \mathcal{D}^1 has one of the properties **A1** or **A3**. If \mathcal{D}^2 D -dominates \mathcal{D}^1 (with a constant B) then \mathcal{D}^2 has the same property as \mathcal{D}^1 : **A1** with $C_1^2 = C_1^1 B$ or **A3** with $V^2 = V^1 B$.

Proof. In both cases **A1** and **A3** the proof is the same. We demonstrate the case **A3**. Let $f = \sum_{i \in T} x_i g_i^2$. Then by the **A3** property of \mathcal{D}^1 we have

$$\sum_{i \in A} |x_i| \leq V^1 |A|^r \left\| \sum_{i \in A} x_i g_i^1 - \sum_{i \in \Lambda} c_i g_i^1 \right\| \leq V^1 B |A|^r \left\| \sum_{i \in A} x_i g_i^2 - \sum_{i \in \Lambda} c_i g_i^2 \right\|.$$

□

Proposition 2.2. *Assume \mathcal{D}^1 has the property **A2**. If \mathcal{D}^1 and \mathcal{D}^2 are D -equivalent (with constants E_1 and E_2) then \mathcal{D}^2 has property **A2** with $U^2 = U^1 E_2 / E_1$.*

Proof. Let $f = \sum_{i \in T} x_i g_i^2$. Then by $\mathcal{D}^1 \approx \mathcal{D}^2$ and the **A2** property of \mathcal{D}^1 we have

$$\begin{aligned} \left\| \sum_{i \in A} x_i g_i^2 - \sum_{i \in \Lambda} c_i g_i^2 \right\| &\geq E_1 \left\| \sum_{i \in A} x_i g_i^1 - \sum_{i \in \Lambda} c_i g_i^1 \right\| \\ &\geq E_1 (U^1)^{-1} \left\| \sum_{i \in A} x_i g_i^1 \right\| \geq (E_1 / E_2) (U^1)^{-1} \left\| \sum_{i \in A} x_i g_i^2 \right\|. \end{aligned}$$

□

We now proceed to main results of [18] and [41] on the WCGA with respect to redundant dictionaries. The following Theorem 2.1 from [41] in the case $q = 2$ was proved in [18].

Theorem 2.1. *Let X be a Banach space with $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. Suppose K -sparse f^ϵ satisfies **A1**, **A2** and $\|f_0 - f^\epsilon\| \leq \epsilon$. Assume that $r q' \geq 1$. Then the WCGA with weakness parameter t applied to f_0 provides*

$$\|f_{C(t, \gamma, C_1) U^{q'} \ln(U+1) K^{r q'}}\| \leq C \epsilon \quad \text{for} \quad K + C(t, \gamma, C_1) U^{q'} \ln(U+1) K^{r q'} \leq D$$

with an absolute constant C .

It was pointed out in [18] that Theorem 2.1 provides a corollary for Hilbert spaces that gives sufficient conditions somewhat weaker than the known RIP conditions on \mathcal{D} for the Lebesgue-type inequality to hold. We formulate the corresponding definitions and results. Let \mathcal{D} be the Riesz dictionary with depth D and parameter $\delta \in (0, 1)$. This class of dictionaries is a generalization of the class of classical Riesz bases. We give a definition in a general Hilbert space (see [40], p. 306).

Definition 2.2. *A dictionary \mathcal{D} is called the Riesz dictionary with depth D and parameter $\delta \in (0, 1)$ if, for any D distinct elements e_1, \dots, e_D of the dictionary and any coefficients $a = (a_1, \dots, a_D)$, we have*

$$(1 - \delta) \|a\|_2^2 \leq \left\| \sum_{i=1}^D a_i e_i \right\|^2 \leq (1 + \delta) \|a\|_2^2. \quad (2.5)$$

We denote the class of Riesz dictionaries with depth D and parameter $\delta \in (0, 1)$ by $R(D, \delta)$.

The term Riesz dictionary with depth D and parameter $\delta \in (0, 1)$ is another name for a dictionary satisfying the Restricted Isometry Property (RIP) with parameters D and δ . The following simple lemma holds.

Lemma 2.1. *Let $\mathcal{D} \in R(D, \delta)$ and let $e_j \in \mathcal{D}$, $j = 1, \dots, s$. For $f = \sum_{i=1}^s a_i e_i$ and $A \subset \{1, \dots, s\}$ denote*

$$S_A(f) := \sum_{i \in A} a_i e_i.$$

If $s \leq D$ then

$$\|S_A(f)\|^2 \leq (1 + \delta)(1 - \delta)^{-1} \|f\|^2.$$

Lemma 2.1 implies that if $\mathcal{D} \in R(D, \delta)$ then it is (D, D) -unconditional with a constant $U = (1 + \delta)^{1/2}(1 - \delta)^{-1/2}$.

Theorem 2.2. *Let X be a Hilbert space. Suppose K -sparse f^ϵ satisfies **A2** and $\|f_0 - f^\epsilon\| \leq \epsilon$. Then the WOMP with weakness parameter t applied to f_0 provides*

$$\|f_{C(t,U)K}\| \leq C\epsilon \quad \text{for} \quad K + C(t, U)K \leq D$$

with an absolute constant C .

Theorem 2.2 implies the following corollaries.

Corollary 2.1. *Let X be a Hilbert space. Suppose any K -sparse f satisfies **A2**. Then the WOMP with weakness parameter t applied to f_0 provides*

$$\|f_{C(t,U)K}\| \leq C\sigma_K(f_0, \mathcal{D}) \quad \text{for} \quad K + C(t, U)K \leq D$$

with an absolute constant C .

Corollary 2.2. *Let X be a Hilbert space. Suppose $\mathcal{D} \in R(D, \delta)$. Then the WOMP with weakness parameter t applied to f_0 provides*

$$\|f_{C(t,\delta)K}\| \leq C\sigma_K(f_0, \mathcal{D}) \quad \text{for} \quad K + C(t, \delta)K \leq D$$

with an absolute constant C .

We emphasized in [18] that in Theorem 2.1 we impose our conditions on an individual function f^ϵ . It may happen that the dictionary does not have the Nikol'skii $\ell_1 X$ property and (K, D) -unconditionality but the given f_0 can

be approximated by f^ϵ which does satisfy assumptions **A1** and **A2**. Even in the case of a Hilbert space the above results from [18] add something new to the study based on the RIP property of a dictionary. First of all, Theorem 2.2 shows that it is sufficient to impose assumption **A2** on f^ϵ in order to obtain exact recovery and the Lebesgue-type inequality results. Second, Corollary 2.1 shows that the condition **A2**, which is weaker than the RIP condition, is sufficient for exact recovery and the Lebesgue-type inequality results. Third, Corollary 2.2 shows that even if we impose our assumptions in terms of RIP we do not need to assume that $\delta < \delta_0$. In fact, the result works for all $\delta < 1$ with parameters depending on δ .

Theorem 2.1 follows from the combination of Theorems 2.3 and 2.4. In case $q = 2$ these theorems were proved in [18] and in general case $q \in (1, 2]$ – in [41].

Theorem 2.3. *Let X be a Banach space with $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. Suppose for a given f_0 we have $\|f_0 - f^\epsilon\| \leq \epsilon$ with K -sparse $f := f^\epsilon$ satisfying **A3**. Then for any $k \geq 0$ we have for $K + m \leq D$*

$$\|f_m\| \leq \|f_k\| \exp\left(-\frac{c_1(m-k)}{K^{rq'}}\right) + 2\epsilon, \quad q' := \frac{q}{q-1},$$

$$\text{where } c_1 := \frac{t^{q'}}{2(16\gamma)^{\frac{1}{q-1}} V^{q'}}.$$

In all theorems that follow we assume $rq' \geq 1$.

Theorem 2.4. *Let X be a Banach space with $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. Suppose K -sparse f^ϵ satisfies **A1**, **A2** and $\|f_0 - f^\epsilon\| \leq \epsilon$. Then the WCGA with weakness parameter t applied to f_0 provides*

$$\|f_{C'U^{q'} \ln(U+1)K^{rq'}}\| \leq CU\epsilon \quad \text{for} \quad K + C'U^{q'} \ln(U+1)K^{rq'} \leq D$$

with an absolute constant C and $C' = C_2(q)\gamma^{\frac{1}{q-1}}C_1^{q'}t^{-q'}$.

We formulate an immediate corollary of Theorem 2.4 with $\epsilon = 0$.

Corollary 2.3. *Let X be a Banach space with $\rho(u) \leq \gamma u^q$. Suppose K -sparse f satisfies **A1**, **A2**. Then the WCGA with weakness parameter t applied to f recovers it exactly after $C'U^{q'} \ln(U+1)K^{rq'}$ iterations under condition $K + C'U^{q'} \ln(U+1)K^{rq'} \leq D$.*

We formulate versions of Theorem 2.4 with assumptions **A1**, **A2** replaced by a single assumption **A3** and replaced by two assumptions **A2** and **A3**. The corresponding modifications in the proofs go as in the proof of Theorem 2.3.

Theorem 2.5. *Let X be a Banach space with $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. Suppose K -sparse f^ϵ satisfies **A3** and $\|f_0 - f^\epsilon\| \leq \epsilon$. Then the WCGA with weakness parameter t applied to f_0 provides*

$$\|f_{C(t,\gamma,q)V^{q'}\ln(VK)K^{rq'}}\| \leq CVK^r\epsilon \quad \text{for} \quad K + C(t,\gamma,q)V^{q'}\ln(VK)K^{rq'} \leq D$$

with an absolute constant C and $C(t,\gamma,q) = C_2(q)\gamma^{\frac{1}{q-1}}t^{-q'}$.

Theorem 2.6. *Let X be a Banach space with $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. Suppose K -sparse f^ϵ satisfies **A2**, **A3** and $\|f_0 - f^\epsilon\| \leq \epsilon$. Then the WCGA with weakness parameter t applied to f_0 provides*

$$\|f_{C(t,\gamma,q)V^{q'}\ln(U+1)K^{rq'}}\| \leq CU\epsilon \quad \text{for} \quad K + C(t,\gamma,q)V^{q'}\ln(U+1)K^{rq'} \leq D$$

with an absolute constant C and $C(t,\gamma,q) = C_2(q)\gamma^{\frac{1}{q-1}}t^{-q'}$.

Theorems 2.5 and 2.3 imply the following analog of Theorem 2.1.

Theorem 2.7. *Let X be a Banach space with $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. Suppose K -sparse f^ϵ satisfies **A3** and $\|f_0 - f^\epsilon\| \leq \epsilon$. Then the WCGA with weakness parameter t applied to f_0 provides*

$$\|f_{C(t,\gamma,q)V^{q'}\ln(VK)K^{rq'}}\| \leq C\epsilon \quad \text{for} \quad K + C(t,\gamma,q)V^{q'}\ln(VK)K^{rq'} \leq D$$

with an absolute constant C and $C(t,\gamma,q) = C_2(q)\gamma^{\frac{1}{q-1}}t^{-q'}$.

The following edition of Theorems 2.1 and 2.7 is also useful in applications. It follows from Theorems 2.6 and 2.3.

Theorem 2.8. *Let X be a Banach space with $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. Suppose K -sparse f^ϵ satisfies **A2**, **A3** and $\|f_0 - f^\epsilon\| \leq \epsilon$. Then the WCGA with weakness parameter t applied to f_0 provides*

$$\|f_{C(t,\gamma,q)V^{q'}\ln(U+1)K^{rq'}}\| \leq C\epsilon \quad \text{for} \quad K + C(t,\gamma,q)V^{q'}\ln(U+1)K^{rq'} \leq D$$

with an absolute constant C and $C(t,\gamma,q) = C_2(q)\gamma^{\frac{1}{q-1}}t^{-q'}$.

3 Proofs

Proof of Theorem 2.3. We begin with a proof of Theorem 2.3.

Proof. Let

$$f := f^\epsilon = \sum_{i \in T} x_i g_i, \quad |T| = K, \quad g_i \in \mathcal{D}.$$

Denote by T^m the set of indices of $g_j \in \mathcal{D}$, $j \in T$, picked by the WCGA after m iterations, $\Gamma^m := T \setminus T^m$. Denote by $A_1(\mathcal{D})$ the closure in X of the convex hull of the symmetrized dictionary $\mathcal{D}^\pm := \{\pm g, g \in \mathcal{D}\}$. We will bound $\|f_m\|$ from above. Assume $\|f_{m-1}\| \geq \epsilon$. Let $m > k$. We bound from below

$$S_m := \sup_{\phi \in A_1(\mathcal{D})} |F_{f_{m-1}}(\phi)|.$$

Denote $A_m := \Gamma^{m-1}$. Then

$$S_m \geq F_{f_{m-1}}(f_{A_m} / \|f_{A_m}\|_1),$$

where $\|f_A\|_1 := \sum_{i \in A} |x_i|$. Next, by Lemma 6.9, p. 342, from [40] we obtain

$$F_{f_{m-1}}(f_{A_m}) = F_{f_{m-1}}(f^\epsilon) \geq \|f_{m-1}\| - \epsilon.$$

Thus

$$S_m \geq \|f_{A_m}\|_1^{-1} (\|f_{m-1}\| - \epsilon). \quad (3.1)$$

From the definition of the modulus of smoothness we have for any λ

$$\|f_{m-1} - \lambda \varphi_m\| + \|f_{m-1} + \lambda \varphi_m\| \leq 2\|f_{m-1}\| \left(1 + \rho \left(\frac{\lambda}{\|f_{m-1}\|} \right) \right) \quad (3.2)$$

and by (1) from the definition of the WCGA and Lemma 6.10 from [40], p. 343, we get

$$\begin{aligned} |F_{f_{m-1}}(\varphi_m)| &\geq t \sup_{g \in \mathcal{D}} |F_{f_{m-1}}(g)| = \\ &t \sup_{\phi \in A_1(\mathcal{D})} |F_{f_{m-1}}(\phi)| = tS_m. \end{aligned}$$

Then either $F_{f_{m-1}}(\varphi_m) \geq tS_m$ or $F_{f_{m-1}}(-\varphi_m) \geq tS_m$. Both cases are treated in the same way. We demonstrate the case $F_{f_{m-1}}(\varphi_m) \geq tS_m$. We have for $\lambda \geq 0$

$$\|f_{m-1} + \lambda \varphi_m\| \geq F_{f_{m-1}}(f_{m-1} + \lambda \varphi_m) \geq \|f_{m-1}\| + \lambda tS_m.$$

From here and from (3.2) we obtain

$$\|f_m\| \leq \|f_{m-1} - \lambda \varphi_m\| \leq \|f_{m-1}\| + \inf_{\lambda \geq 0} (-\lambda t S_m + 2\|f_{m-1}\| \rho(\lambda/\|f_{m-1}\|)).$$

We discuss here the case $\rho(u) \leq \gamma u^q$. Using (3.1) we get

$$\|f_m\| \leq \|f_{m-1}\| \left(1 - \frac{\lambda t}{\|f_{A_m}\|_1} + 2\gamma \frac{\lambda^q}{\|f_{m-1}\|^q} \right) + \frac{\epsilon \lambda t}{\|f_{A_m}\|_1}.$$

Let λ_1 be a solution of

$$\frac{\lambda t}{2\|f_{A_m}\|_1} = 2\gamma \frac{\lambda^q}{\|f_{m-1}\|^q}, \quad \lambda_1 = \left(\frac{t\|f_{m-1}\|^q}{4\gamma\|f_{A_m}\|_1} \right)^{\frac{1}{q-1}}.$$

Our assumption (2.4) gives

$$\begin{aligned} \|f_{A_m}\|_1 &= \|(f^\epsilon - G_{m-1})_{A_m}\|_1 \leq V K^r \|f^\epsilon - G_{m-1}\| \\ &\leq V K^r (\|f_0 - G_{m-1}\| + \|f_0 - f^\epsilon\|) \leq V K^r (\|f_{m-1}\| + \epsilon). \end{aligned} \quad (3.3)$$

Specify

$$\lambda = \left(\frac{t\|f_{A_m}\|_1^{q-1}}{16\gamma(VK^r)^q} \right)^{\frac{1}{q-1}}.$$

Then, using $\|f_{m-1}\| \geq \epsilon$ we get

$$\left(\frac{\lambda}{\lambda_1} \right)^{q-1} = \frac{\|f_{A_m}\|_1^q}{4\|f_{m-1}\|^q (VK^r)^q} \leq 1$$

and obtain

$$\|f_m\| \leq \|f_{m-1}\| \left(1 - \frac{t^{q'}}{2(16\gamma)^{\frac{1}{q-1}} (VK^r)^{q'}} \right) + \frac{\epsilon t^{q'}}{(16\gamma)^{\frac{1}{q-1}} (VK^r)^{q'}}. \quad (3.4)$$

Denote $c_1 := \frac{t^{q'}}{2(16\gamma)^{\frac{1}{q-1}} V^{q'}}$. Then

$$\|f_m\| \leq \|f_k\| \exp \left(-\frac{c_1(m-k)}{K^{rq'}} \right) + 2\epsilon.$$

□

Proof of Theorem 2.4. We proceed to a proof of Theorem 2.4. Modifications of this proof which are in a style of the above proof of Theorem 2.3 give Theorems 2.5 and 2.6.

Proof. We begin with a brief description of the structure of the proof. We are given f_0 and $f := f^\epsilon$ such that $\|f_0 - f\| \leq \epsilon$ and f is K -sparse satisfying **A1** and **A2**. We apply the WCGA to f_0 and control how many dictionary elements g_i from the representation of f

$$f := f^\epsilon := \sum_{i \in T} x_i g_i$$

are picked up by the WCGA after m iterations. As above denote by T^m the set of indices $i \in T$ such that g_i has been taken by the WCGA at one of the first m iterations. Denote $\Gamma^m := T \setminus T^m$. It is clear that if $\Gamma^m = \emptyset$ then $\|f_m\| \leq \epsilon$ because in this case $f \in \Phi_m$.

Our analysis goes as follows. For a residual f_k we assume that Γ^k is nonempty. Then we prove that after $N(k)$ iterations we arrive at a residual $f_{k'}$, $k' = k + N(k)$, such that either

$$\|f_{k'}\| \leq CU\epsilon \quad (3.5)$$

or

$$|\Gamma^{k'}| < |\Gamma^k| - 2^{L-2} \quad (3.6)$$

with some natural number L . An important fact is that for the number $N(k)$ of iterations we have a bound

$$N(k) \leq \beta 2^{aL}, \quad a := rq'. \quad (3.7)$$

Next, we prove that if we begin with $k = 0$ and apply the above argument to the sequence of residuals $f_0, f_{k_1}, \dots, f_{k_s}$, then after not more than $N := 2^{2a+1}\beta K^a$ iterations, we obtain either $\|f_N\| \leq CU\epsilon$ or $\Gamma^N = \emptyset$, which in turn implies that $\|f_N\| \leq \epsilon$.

We now proceed to the detailed argument. The following corollary of (2.3) will be often used: for $m \leq D - K$ and $A \subset \Gamma^m$ we have

$$\|f_A\| \leq U(\|f_m\| + \epsilon). \quad (3.8)$$

It follows from the fact that $f_A - f + G_m$ has the form $\sum_{i \in \Lambda} c_i g_i$ with Λ satisfying $|A| + |\Lambda| \leq D$, $A \cap \Lambda = \emptyset$, and from our assumption $\|f - f_0\| \leq \epsilon$.

The following lemma plays a key role in the proof.

Lemma 3.1. *Let f satisfy **A1** and **A2** and let $A \subset \Gamma^k$ be nonempty. Denote $B := \Gamma^k \setminus A$. Then for any $m \in (k, D - K]$ we have either $\|f_{m-1}\| \leq \epsilon$ or*

$$\|f_m\| \leq \|f_{m-1}\|(1 - u) + 2u(\|f_B\| + \epsilon), \quad (3.9)$$

where

$$u := c_1|A|^{-rq'}, \quad c_1 := \frac{t^{q'}}{2(16\gamma)^{\frac{1}{q-1}}(C_1U)^{q'}},$$

with C_1 and U from **A1** and **A2**.

Proof. As above in the proof of Theorem 2.3 we bound S_m from below. It is clear that $S_m \geq 0$. Denote $A(m) := A \cap \Gamma^{m-1}$. Then

$$S_m \geq F_{f_{m-1}}(f_{A(m)})/\|f_{A(m)}\|_1.$$

Next,

$$F_{f_{m-1}}(f_{A(m)}) = F_{f_{m-1}}(f_{A(m)} + f_B - f_B).$$

Then $f_{A(m)} + f_B = f^\epsilon - f_\Lambda$ with $F_{f_{m-1}}(f_\Lambda) = 0$. Moreover, it is easy to see that $F_{f_{m-1}}(f^\epsilon) \geq \|f_{m-1}\| - \epsilon$. Therefore,

$$F_{f_{m-1}}(f_{A(m)} + f_B - f_B) \geq \|f_{m-1}\| - \epsilon - \|f_B\|.$$

Thus

$$S_m \geq \|f_{A(m)}\|_1^{-1} \max(0, \|f_{m-1}\| - \epsilon - \|f_B\|).$$

By (2.2) we get

$$\|f_{A(m)}\|_1 \leq C_1|A(m)|^r\|f_{A(m)}\| \leq C_1|A|^r\|f_{A(m)}\|.$$

Then

$$S_m \geq \frac{\|f_{m-1}\| - \|f_B\| - \epsilon}{C_1|A|^r\|f_{A(m)}\|}. \quad (3.10)$$

From the definition of the modulus of smoothness we have for any λ

$$\|f_{m-1} - \lambda\varphi_m\| + \|f_{m-1} + \lambda\varphi_m\| \leq 2\|f_{m-1}\| \left(1 + \rho\left(\frac{\lambda}{\|f_{m-1}\|}\right)\right)$$

and by (1) from the definition of the WCGA and Lemma 6.10 from [40], p. 343, we get

$$|F_{f_{m-1}}(\varphi_m)| \geq t \sup_{g \in \mathcal{D}} |F_{f_{m-1}}(g)| =$$

$$t \sup_{\phi \in A_1(\mathcal{D})} |F_{f_{m-1}}(\phi)|.$$

From here we obtain

$$\|f_m\| \leq \|f_{m-1}\| + \inf_{\lambda \geq 0} (-\lambda t S_m + 2\|f_{m-1}\| \rho(\lambda/\|f_{m-1}\|)).$$

We discuss here the case $\rho(u) \leq \gamma u^q$. Using (3.10) we get for any $\lambda \geq 0$

$$\|f_m\| \leq \|f_{m-1}\| \left(1 - \frac{\lambda t}{C_1 |A|^r \|f_{A(m)}\|} + 2\gamma \frac{\lambda^q}{\|f_{m-1}\|^q} \right) + \frac{\lambda t (\|f_B\| + \epsilon)}{C_1 |A|^r \|f_{A(m)}\|}.$$

Let λ_1 be a solution of

$$\frac{\lambda t}{2C_1 |A|^r \|f_{A(m)}\|} = 2\gamma \frac{\lambda^q}{\|f_{m-1}\|^q}, \quad \lambda_1 = \left(\frac{t \|f_{m-1}\|^q}{4\gamma C_1 |A|^r \|f_{A(m)}\|} \right)^{\frac{1}{q-1}}.$$

Inequality (3.8) gives

$$\|f_{A(m)}\| \leq U(\|f_{m-1}\| + \epsilon).$$

Specify

$$\lambda = \left(\frac{t \|f_{A(m)}\|^{q-1}}{16\gamma C_1 |A|^r U^q} \right)^{\frac{1}{q-1}}.$$

Then $\lambda \leq \lambda_1$ and we obtain

$$\|f_m\| \leq \|f_{m-1}\| \left(1 - \frac{t^{q'}}{2(16\gamma)^{\frac{1}{q-1}} (C_1 U |A|^r)^{q'}} \right) + \frac{t^{q'} (\|f_B\| + \epsilon)}{(16\gamma)^{\frac{1}{q-1}} (C_1 |A|^r U)^{q'}}. \quad (3.11)$$

□

For simplicity of notations we consider separately the case $|\Gamma^k| \geq 2$ and the case $|\Gamma^k| = 1$. We begin with the generic case $|\Gamma^k| \geq 2$. We apply Lemma 3.1 with different pairs A_j, B_j , which we now construct. Let n be a natural number such that

$$2^{n-1} < |\Gamma^k| \leq 2^n.$$

For $j = 1, 2, \dots, n, n+1$ consider the following pairs of sets A_j, B_j : $A_{n+1} = \Gamma^k$, $B_{n+1} = \emptyset$; for $j \leq n$, $A_j := \Gamma^k \setminus B_j$ with $B_j \subset \Gamma^k$ is such that $|B_j| \geq |\Gamma^k| - 2^{j-1}$ and for any set $J \subset \Gamma^k$ with $|J| \geq |\Gamma^k| - 2^{j-1}$ we have

$$\|f_{B_j}\| \leq \|f_J\|.$$

We note that the above definition implies that $|A_j| \leq 2^{j-1}$ and that if for some $Q \subset \Gamma^k$ we have

$$\|f_Q\| < \|f_{B_j}\| \quad \text{then} \quad |Q| < |\Gamma^k| - 2^{j-1}. \quad (3.12)$$

Set $B_0 := \Gamma^k$. Note that property (3.12) is obvious for $j = 0$.

Let $j_0 \in [1, n]$ be an index such that if $j_0 = 1$ then $B_1 \neq \Gamma^k$ and if $j_0 \geq 2$ then

$$B_1 = B_2 = \dots = B_{j_0-1} = \Gamma^k, \quad B_{j_0} \neq \Gamma^k.$$

For a given $b > 1$, to be specified later, denote by $L := L(b)$ the index such that $(B_0 := \Gamma^k)$

$$\begin{aligned} \|f_{B_0}\| &< b\|f_{B_{j_0}}\|, \\ \|f_{B_{j_0}}\| &< b\|f_{B_{j_0+1}}\|, \\ &\dots \\ \|f_{B_{L-2}}\| &< b\|f_{B_{L-1}}\|, \\ \|f_{B_{L-1}}\| &\geq b\|f_{B_L}\|. \end{aligned}$$

Then

$$\|f_{B_j}\| \leq b^{L-1-j}\|f_{B_{L-1}}\|, \quad j = j_0, \dots, L, \quad (3.13)$$

and

$$\|f_{B_0}\| = \dots = \|f_{B_{j_0-1}}\| \leq b^{L-j_0}\|f_{B_{L-1}}\|. \quad (3.14)$$

Clearly, $L \leq n + 1$.

Define $m_0 := \dots m_{j_0-1} := k$ and, inductively,

$$m_j = m_{j-1} + [\beta|A_j|^{rq'}], \quad j = j_0, \dots, L,$$

where $[x]$ denotes the integer part of x . The parameter β is any, which satisfies the following inequalities with c_1 from Lemma 3.1

$$\beta \geq 1, \quad e^{-c_1\beta/2} < 1/2, \quad 16Ue^{-c_1\beta/2} < 1. \quad (3.15)$$

We note that the inequality $\beta \geq 1$ implies that

$$[\beta|A_j|^{rq'}] \geq \beta|A_j|^{rq'}/2.$$

Taking into account that $rq' \geq 1$ and $|A_j| \geq 1$ we obtain

$$m_j \geq m_{j-1} + 1.$$

At iterations from $m_{j-1} + 1$ to m_j we apply Lemma 3.1 with $A = A_j$ and obtain from (3.9) that either $\|f_{m-1}\| \leq \epsilon$ or

$$\|f_m\| \leq \|f_{m-1}\|(1-u) + 2u(\|f_{B_j}\| + \epsilon), \quad u := c_1|A_j|^{-rq'}.$$

Using $1-u \leq e^{-u}$ and $\sum_{k=0}^{\infty}(1-u)^k = 1/u$ we derive from here

$$\|f_{m_j}\| \leq \|f_{m_{j-1}}\|e^{-c_1\beta/2} + 2(\|f_{B_j}\| + \epsilon). \quad (3.16)$$

We continue it up to $j = L$. Denote $\eta := e^{-c_1\beta/2}$. Then either $\|f_{m_L}\| \leq \epsilon$ or

$$\|f_{m_L}\| \leq \|f_k\|\eta^{L-j_0+1} + 2\sum_{j=j_0}^L(\|f_{B_j}\| + \epsilon)\eta^{L-j}.$$

We bound the $\|f_k\|$. It follows from the definition of f_k that $\|f_k\|$ is the error of best approximation of f_0 by the subspace Φ_k . Representing $f_0 = f + f_0 - f$ we see that $\|f_k\|$ is not greater than the error of best approximation of f by the subspace Φ_k plus $\|f_0 - f\|$. This implies $\|f_k\| \leq \|f_{B_0}\| + \epsilon$. Therefore we continue

$$\begin{aligned} &\leq (\|f_{B_0}\| + \epsilon)\eta^{L-j_0+1} + 2\sum_{j=j_0}^L(\|f_{B_{L-1}}\|(\eta b)^{L-j}b^{-1} + \epsilon\eta^{L-j}) \\ &\leq b^{-1}\|f_{B_{L-1}}\| \left((\eta b)^{L-j_0+1} + 2\sum_{j=j_0}^L(\eta b)^{L-j} \right) + \frac{2\epsilon}{1-\eta}. \end{aligned}$$

Our choice of β guarantees $\eta < 1/2$. Choose $b = \frac{1}{2\eta}$. Then

$$\|f_{m_L}\| \leq \|f_{B_{L-1}}\|8e^{-c_1\beta/2} + 4\epsilon. \quad (3.17)$$

By (3.8) we get

$$\|f_{\Gamma^{m_L}}\| \leq U(\|f_{m_L}\| + \epsilon) \leq U(\|f_{B_{L-1}}\|8e^{-c_1\beta/2} + 5\epsilon).$$

If $\|f_{B_{L-1}}\| \leq 10U\epsilon$ then by (3.17)

$$\|f_{m_L}\| \leq CU\epsilon, \quad C = 44.$$

If $\|f_{B_{L-1}}\| \geq 10U\epsilon$ then by our choice of β we have $16Ue^{-c_1\beta/2} < 1$ and

$$U(\|f_{B_{L-1}}\|8e^{-c_1\beta/2} + 5\epsilon) < \|f_{B_{L-1}}\|.$$

Therefore

$$\|f_{\Gamma^{m_L}}\| < \|f_{B_{L-1}}\|.$$

This implies

$$|\Gamma^{m_L}| < |\Gamma^k| - 2^{L-2}.$$

In the above proof our assumption $j_0 \leq n$ is equivalent to the assumption that $B_n \neq \Gamma^k$. We now consider the case $B_n = \Gamma^k$ and, therefore, $B_j = \Gamma^k$, $j = 0, 1, \dots, n$. This means that $\|f_{\Gamma^k}\| \leq \|f_J\|$ for any J with $|J| \geq |\Gamma^k| - 2^{n-1}$. Therefore, if for some $Q \subset \Gamma^k$ we have

$$\|f_Q\| < \|f_{\Gamma^k}\| \quad \text{then} \quad |Q| < |\Gamma^k| - 2^{n-1}. \quad (3.18)$$

In this case we set $m_0 := k$ and

$$m_1 := k + [\beta |\Gamma^k|^{r_{q'}}].$$

Then by Lemma 3.1 with $A = \Gamma^k$ we obtain as in (3.16)

$$\|f_{m_1}\| \leq \|f_{m_0}\| e^{-c_1 \beta/2} + 2\epsilon \leq \|f_{\Gamma^k}\| e^{-c_1 \beta/2} + 3\epsilon. \quad (3.19)$$

By (3.8) we get

$$\|f_{\Gamma^{m_1}}\| \leq U(\|f_{m_1}\| + \epsilon) \leq U(\|f_{\Gamma^k}\| e^{-c_1 \beta/2} + 4\epsilon).$$

If $\|f_{\Gamma^k}\| \leq 8U\epsilon$ then by (3.19)

$$\|f_{m_1}\| \leq 7U\epsilon.$$

If $\|f_{\Gamma^k}\| \geq 8U\epsilon$ then by our choice of β we have $2Ue^{-c_1 \beta/2} < 1$ and

$$\|f_{\Gamma^{m_1}}\| \leq U(\|f_{\Gamma^k}\| e^{-c_1 \beta/2} + 4\epsilon) < \|f_{\Gamma^k}\|. \quad (3.20)$$

This implies

$$|\Gamma^{m_1}| < |\Gamma^k| - 2^{n-1}.$$

It remains to consider the case $|\Gamma^k| = 1$. By the above argument, where we used Lemma 3.1 with $A = \Gamma^k$ we obtain (3.20). In the case $|\Gamma^k| = 1$ inequality (3.20) implies $\Gamma^{m_1} = \emptyset$, which completes the proof in this case.

We now complete the proof of Theorem 2.4. We begin with f_0 and apply the above argument (with $k = 0$). As a result we either get the required inequality or we reduce the cardinality of support of f from $|T| = K$ to

$|\Gamma^{m_{L_1}}| < |T| - 2^{L_1-2}$ (the WCGA picks up at least 2^{L_1-2} dictionary elements g_i from the representation of f), $m_{L_1} \leq \beta 2^{aL_1}$, $a := rq'$. We continue the process and build a sequence m_{L_j} such that $m_{L_j} \leq \beta 2^{aL_j}$ and after m_{L_j} iterations we reduce the support by at least 2^{L_j-2} . We also note that $m_{L_j} \leq \beta 2^{2a} K^a$. We continue this process till the following inequality is satisfied for the first time

$$m_{L_1} + \dots + m_{L_s} \geq 2^{2a} \beta K^a. \quad (3.21)$$

Then, clearly,

$$m_{L_1} + \dots + m_{L_s} \leq 2^{2a+1} \beta K^a.$$

Using the inequality

$$(a_1 + \dots + a_s)^\theta \leq a_1^\theta + \dots + a_s^\theta, \quad a_j \geq 0, \quad \theta \in (0, 1]$$

we derive from (3.21)

$$\begin{aligned} 2^{L_1-2} + \dots + 2^{L_s-2} &\geq \left(2^{a(L_1-2)} + \dots + 2^{a(L_s-2)}\right)^{\frac{1}{a}} \\ &\geq 2^{-2} \left(2^{aL_1} + \dots + 2^{aL_s}\right)^{\frac{1}{a}} \\ &\geq 2^{-2} \left((\beta)^{-1}(m_{L_1} + \dots + m_{L_s})\right)^{\frac{1}{a}} \geq K. \end{aligned}$$

Thus, after not more than $N := 2^{2a+1} \beta K^a$ iterations we either get the required inequality or we recover f exactly (the WCGA picks up all the dictionary elements g_i from the representation of f) and then $\|f_N\| \leq \|f_0 - f\| \leq \epsilon$. \square

Proof of Theorem 2.5. We begin with a version of Lemma 3.1 that is used in this proof.

Lemma 3.2. *Let f satisfy **A3** and let $A \subset \Gamma^k$ be nonempty. Denote $B := \Gamma^k \setminus A$. Then for any $m \in (k, D - K]$ we have either $\|f_{m-1}\| \leq \epsilon$ or*

$$\|f_m\| \leq \|f_{m-1}\|(1 - u) + 2u(\|f_B\| + \epsilon), \quad (3.22)$$

where

$$u := c_2 |A|^{-rq'}, \quad c_2 := \frac{t^{q'}}{2(16\gamma)^{\frac{1}{q-1}} V^{q'}},$$

with r and V from **A3**.

Proof. The proof is a combination of proofs of Theorem 2.3 and Lemma 3.1. As in the proof of Lemma 3.1 we denote $A(m) := A \cap \Gamma^{m-1}$ and get

$$S_m \geq \|f_{A(m)}\|_1^{-1} \max(0, \|f_{m-1}\| - \epsilon - \|f_B\|).$$

From here in the same way as in the proof of Theorem 2.3 we obtain for any $\lambda \geq 0$

$$\|f_m\| \leq \|f_{m-1}\| \left(1 - \frac{\lambda t}{\|f_{A(m)}\|_1} + 2\gamma \frac{\lambda^q}{\|f_{m-1}\|^q} \right) + \frac{\lambda t(\|f_B\| + \epsilon)}{\|f_{A(m)}\|_1}.$$

Using definition of $A(m)$ we bound by **A3**

$$\begin{aligned} \|f_{A(m)}\|_1 &= \sum_{i \in A(m)} |x_i| \leq V|A(m)|^r \|f_{A(m)} + f - f_{A(m)} - G_{m-1}\| \\ &\leq V|A|^r \|f - G_{m-1}\| \leq V|A|^r (\|f_{m-1}\| + \epsilon). \end{aligned}$$

This inequality is a variant of inequality (3.3) with K replaced by $|A|$. Arguing as in the proof of Theorem 2.3 with K replaced by $|A|$ we obtain the required inequality, which is the corresponding modification (K is replaced by $|A|$ and ϵ is replaced by $\|f_B\| + \epsilon$) of (3.4).

The rest of the proof repeats the proof of Theorem 2.4 with the use of Lemma 3.2 instead of Lemma 3.1 and with the use of the fact that **A3** implies **A2** with $U = VK^r \leq VK$. \square

Proof of Theorem 2.6. This proof repeats the proof of Theorem 2.4 with the use of Lemma 3.2 instead of Lemma 3.1.

4 Examples

In this section, following [41], we discuss applications of Theorems from Section 2 for specific dictionaries \mathcal{D} . Mostly, \mathcal{D} will be a basis Ψ for X . Because of that we use m instead of K in the notation of sparse approximation. In some of our examples we take $X = L_p$, $2 \leq p < \infty$. Then it is known that $\rho(u) \leq \gamma u^2$ with $\gamma = (p-1)/2$. In some other examples we take $X = L_p$, $1 < p \leq 2$. Then it is known that $\rho(u) \leq \gamma u^p$, with $\gamma = 1/p$.

Proposition 4.1. *Let Ψ be a uniformly bounded orthogonal system normalized in $L_p(\Omega)$, $2 \leq p < \infty$, Ω is a bounded domain. Then we have*

$$\|f_{C(t,p,\Omega)m \ln(m+1)}\|_p \leq C \sigma_m(f_0, \Psi)_p. \quad (4.1)$$

The proof of Proposition 4.1 is based on Theorem 2.7.

Corollary 4.1. *Let Ψ be the normalized in L_p , $2 \leq p < \infty$, real d -variate trigonometric system. Then Proposition 4.1 applies and gives for any $f_0 \in L_p$*

$$\|f_{C(t,p,d)m \ln(m+1)}\|_p \leq C\sigma_m(f_0, \Psi)_p. \quad (4.2)$$

We note that (4.2) provides some progress in Open Problem 7.1 (p. 91) from [38].

Proposition 4.2. *Let Ψ be a uniformly bounded orthogonal system normalized in $L_p(\Omega)$, $1 < p \leq 2$, Ω is a bounded domain. Then we have*

$$\|f_{C(t,p,\Omega)m^{p'-1} \ln(m+1)}\|_p \leq C\sigma_m(f_0, \Psi)_p. \quad (4.3)$$

The proof of Proposition 4.2 is based on Theorem 2.7.

Corollary 4.2. *Let Ψ be the normalized in L_p , $1 < p \leq 2$, real d -variate trigonometric system. Then Proposition 4.2 applies and gives for any $f_0 \in L_p$*

$$\|f_{C(t,p,d)m^{p'-1} \ln(m+1)}\|_p \leq C\sigma_m(f_0, \Psi)_p. \quad (4.4)$$

Proposition 4.3. *Let Ψ be the normalized in L_p , $2 \leq p < \infty$, multivariate Haar basis $\mathcal{H}_p^d = \mathcal{H}_p \times \cdots \times \mathcal{H}_p$. Then*

$$\|f_{C(t,p,d)m^{2/p'}}\|_p \leq C\sigma_m(f_0, \mathcal{H}_p^d)_p. \quad (4.5)$$

The proof of Proposition 4.3 is based on Theorem 2.4. Inequality (4.5) provides some progress in Open Problem 7.2 (p. 91) from [38] in the case $2 < p < \infty$.

Proposition 4.4. *Let Ψ be the normalized in L_p , $1 < p \leq 2$, univariate Haar basis $\mathcal{H}_p = \{H_{I,p}\}_I$, where $H_{I,p}$ the Haar functions indexed by dyadic intervals of support of $H_{I,p}$ (we index function 1 by $[0, 1]$ and the first Haar function by $[0, 1)$). Then*

$$\|f_{C(t,p)m}\|_p \leq C\sigma_m(f_0, \mathcal{H}_p)_p. \quad (4.6)$$

The proof of Proposition 4.4 is based on Theorem 2.8. Inequality (4.6) solves the Open Problem 7.2 (p. 91) from [38] in the case $1 < p \leq 2$.

Proposition 4.5. *Let X be a Banach space with $\rho(u) \leq \gamma u^2$. Assume that Ψ is a normalized Schauder basis for X . Then*

$$\|f_{C(t,X,\Psi)m^2 \ln m}\| \leq C\sigma_m(f_0, \Psi). \quad (4.7)$$

The proof of Proposition 4.5 is based on Theorem 2.7. We note that the above bound still works if we replace the assumption that Ψ is a Schauder basis by the assumption that a dictionary \mathcal{D} is $(1, D)$ -unconditional with constant U . Then we obtain

$$\|f_{C(t,\gamma,U)K^2 \ln K}\| \leq C\sigma_K(f_0, \Psi), \quad \text{for } K + C(t, \gamma, U)K^2 \ln K \leq D.$$

Proposition 4.6. *Let X be a Banach space with $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. Assume that Ψ is a normalized Schauder basis for X . Then*

$$\|f_{C(t,X,\Psi)m^{q'} \ln m}\| \leq C\sigma_m(f_0, \Psi). \quad (4.8)$$

The proof of Proposition 4.6 is based on Theorem 2.7. We note that the above bound still works if we replace the assumption that Ψ is a Schauder basis by the assumption that a dictionary \mathcal{D} is $(1, D)$ -unconditional with constant U . Then we obtain

$$\|f_{C(t,\gamma,q,U)K^{q'} \ln K}\| \leq C\sigma_K(f_0, \mathcal{D}), \quad \text{for } K + C(t, \gamma, q, U)K^{q'} \ln K \leq D.$$

We now discuss application of general results of Section 2 to quasi-greedy bases. We begin with a brief introduction to the theory of quasi-greedy bases. Let X be an infinite-dimensional separable Banach space with a norm $\|\cdot\| := \|\cdot\|_X$ and let $\Psi := \{\psi_m\}_{m=1}^\infty$ be a normalized basis for X . The concept of quasi-greedy basis was introduced in [17].

Definition 4.1. *The basis Ψ is called quasi-greedy if there exists some constant C such that*

$$\sup_m \|G_m(f, \Psi)\| \leq C\|f\|.$$

Subsequently, Wojtaszczyk [46] proved that these are precisely the bases for which the TGA merely converges, i.e.,

$$\lim_{n \rightarrow \infty} G_n(f) = f.$$

The following lemma is from [8] (see also [10] and [11] for further discussions).

Lemma 4.1. *Let Ψ be a quasi-greedy basis of X . Then for any finite set of indices Λ we have for all $f \in X$*

$$\|S_\Lambda(f, \Psi)\| \leq C \ln(|\Lambda| + 1) \|f\|,$$

where for $f = \sum_{k=1}^{\infty} c_k(f) \psi_k$ we denote $S_\Lambda(f, \Psi) := \sum_{k \in \Lambda} c_k(f) \psi_k$.

We now formulate a result about quasi-greedy bases in L_p spaces. The following theorem is from [45]. We note that in the case $p = 2$ Theorem 4.1 was proved in [46]. Some notations first. For a given element $f \in X$ we consider the expansion

$$f = \sum_{k=1}^{\infty} c_k(f) \psi_k$$

and the decreasing rearrangement of its coefficients

$$|c_{k_1}(f)| \geq |c_{k_2}(f)| \geq \dots$$

Denote

$$a_n(f) := |c_{k_n}(f)|.$$

Theorem 4.1. *Let $\Psi = \{\psi_m\}_{m=1}^{\infty}$ be a quasi-greedy basis of the L_p space, $1 < p < \infty$. Then for each $f \in X$ we have*

$$C_1(p) \sup_n n^{1/p} a_n(f) \leq \|f\|_p \leq C_2(p) \sum_{n=1}^{\infty} n^{-1/2} a_n(f), \quad 2 \leq p < \infty;$$

$$C_3(p) \sup_n n^{1/2} a_n(f) \leq \|f\|_p \leq C_4(p) \sum_{n=1}^{\infty} n^{1/p-1} a_n(f), \quad 1 < p \leq 2.$$

Proposition 4.7. *Let Ψ be a normalized quasi-greedy basis for L_p , $2 \leq p < \infty$. Then*

$$\|f_{C(t,p)m^{2(1-1/p)} \ln(m+1)}\| \leq C \sigma_m(f_0, \Psi). \quad (4.9)$$

The proof of Proposition 4.7 is based on Theorem 2.7.

Proposition 4.8. *Let Ψ be a normalized quasi-greedy basis for L_p , $1 < p \leq 2$. Then*

$$\|f_{C(t,p)m^{p'/2} \ln(m+1)}\| \leq C \sigma_m(f_0, \Psi). \quad (4.10)$$

The proof of Proposition 4.7 is based on Theorem 2.7.

Proposition 4.9. *Let Ψ be a normalized uniformly bounded orthogonal quasi-greedy basis for L_p , $2 \leq p < \infty$ (for existence of such bases see [23]). Then*

$$\|f_{C(t,p,\Psi)m \ln \ln(m+3)}\|_p \leq C\sigma_m(f_0, \Psi)_p. \quad (4.11)$$

The proof of Proposition 4.7 is based on Theorem 2.8.

Proposition 4.10. *Let Ψ be a normalized uniformly bounded orthogonal quasi-greedy basis for L_p , $1 < p \leq 2$ (for existence of such bases see [23]). Then*

$$\|f_{C(t,p,\Psi)m^{p'/2} \ln \ln(m+3)}\|_p \leq C\sigma_m(f_0, \Psi)_p. \quad (4.12)$$

The proof of Proposition 4.7 is based on Theorem 2.8.

Proposition 4.4 is the first result about almost greedy bases with respect to WCGA in Banach spaces. It shows that the univariate Haar basis is an almost greedy basis with respect to the WCGA in the L_p spaces for $1 < p \leq 2$. Proposition 4.1 shows that uniformly bounded orthogonal bases are ϕ -greedy bases with respect to WCGA with $\phi(u) = C(t, p, \Omega) \ln(u+1)$ in the L_p spaces for $2 \leq p < \infty$. We do not know if these bases are almost greedy with respect to WCGA. They are good candidates for that.

It is known (see [40], p. 17) that the univariate Haar basis is a greedy basis with respect to TGA for all L_p , $1 < p < \infty$. Proposition 4.3 only shows that it is a ϕ -greedy basis with respect to WCGA with $\phi(u) = C(t, p)u^{1-2/p}$ in the L_p spaces for $2 \leq p < \infty$. It is much weaker than the corresponding results for the \mathcal{H}_p , $1 < p \leq 2$, and for the trigonometric system, $2 \leq p < \infty$ (see Corollary 4.1). We do not know if this result on the Haar basis can be substantially improved. At the level of our today's technique we can observe that the Haar basis is ideal (greedy basis) for the TGA in L_p , $1 < p < \infty$, almost ideal (almost greedy basis) for the WCGA in L_p , $1 < p \leq 2$, and that the trigonometric system is very good for the WCGA in L_p , $2 \leq p < \infty$.

Corollary 4.2 shows that our results for the trigonometric system in L_p , $1 < p < 2$, are not as strong as for $2 \leq p < \infty$. We do not know if it is a lack of appropriate technique or it reflects the nature of the WCGA with respect to the trigonometric system.

We note that Propositions 2.1 and 2.2 can be used to formulate the above Propositions for a more general bases. In these cases we use Propositions 2.1 and 2.2 with $D = \infty$. In Propositions 4.1, 4.2, 4.7, and 4.8, where we used Theorem 2.7, we can replace the basis Ψ by a basis Φ , which dominates the basis Ψ . In Propositions 4.3, 4.4, 4.9, and 4.10, where we used either Theorem

2.4 or 2.8, we can replace the basis Ψ by a basis Φ , which is equivalent to the basis Ψ .

It is interesting to compare Theorem 2.3 with the following known result. The following theorem provides rate of convergence (see [40], p. 347). We denote by $A_1(\mathcal{D})$ the closure in X of the convex hull of the symmetrized dictionary $\mathcal{D}^\pm := \{\pm g : g \in \mathcal{D}\}$.

Theorem 4.2. *Let X be a uniformly smooth Banach space with modulus of smoothness $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. Take a number $\epsilon \geq 0$ and two elements f_0, f^ϵ from X such that*

$$\|f_0 - f^\epsilon\| \leq \epsilon, \quad f^\epsilon/A(\epsilon) \in A_1(\mathcal{D}),$$

with some number $A(\epsilon) > 0$. Then, for the WCGA we have

$$\|f_m^{c,t}\| \leq \max(2\epsilon, C(q, \gamma)(A(\epsilon) + \epsilon)t(1 + m)^{1/q-1}).$$

Both Theorem 4.2 and Theorem 2.3 provide stability of the WCGA with respect to noise. In order to apply them for noisy data we interpret f_0 as a noisy version of a signal and f^ϵ as a noiseless version of a signal. Then, assumption $f^\epsilon/A(\epsilon) \in A_1(\mathcal{D})$ describes our smoothness assumption on the noiseless signal and assumption $f^\epsilon \in \Sigma_K(\mathcal{D})$ describes our structural assumption on the noiseless signal. In fact, Theorem 4.2 simultaneously takes care of two issues: noisy data and approximation in an interpolation space. Theorem 4.2 can be applied for approximation of f_0 under assumption that f_0 belongs to one of interpolation spaces between X and the space generated by the $A_1(\mathcal{D})$ -norm (atomic norm).

5 Sparse trigonometric approximation

Sparse trigonometric approximation of periodic functions began by the paper of S.B. Stechkin [28], who used it in the criterion for absolute convergence of trigonometric series. R.S. Ismagilov [14] found nontrivial estimates for m -term approximation of functions with singularities of the type $|x|$ and gave interesting and important applications to the widths of Sobolev classes. He used a deterministic method based on number theoretical constructions. His method was developed by V.E. Maiorov [19], who used a method based on Gaussian sums. Further strong results were obtained in [7] with the help of a nonconstructive result from finite dimensional Banach spaces due to E.D.

Gluskin [12]. Other powerful nonconstructive method, which is based on a probabilistic argument, was used by Y. Makovoz [20] and by E.S. Belinskii [2]. Different methods were created in [31], [16], [36], [42] for proving lower bounds for function classes. It was discovered in [9] and [39] that greedy algorithms can be used for constructive m -term approximation with respect to the trigonometric system. We demonstrate in [43] how greedy algorithms can be used to prove optimal or best known upper bounds for m -term approximation of classes of functions with mixed smoothness. It is a simple and powerful method of proving upper bounds. The reader can find a detailed study of m -term approximation of classes of functions with mixed smoothness, including small smoothness, in the paper [24] by A.S. Romanyuk and in [43], [44]. We note that in the case $2 < p < \infty$ the upper bounds in [24] are not constructive.

We discuss some approximation problems for classes of functions with mixed smoothness. We define these classes momentarily. We will begin with the case of univariate periodic functions. Let for $r > 0$

$$F_r(x) := 1 + 2 \sum_{k=1}^{\infty} k^{-r} \cos(kx - r\pi/2) \quad (5.1)$$

and

$$W_p^r := \{f : f = \varphi * F_r, \quad \|\varphi\|_p \leq 1\}. \quad (5.2)$$

It is well known that for $r > 1/p$ the class W_p^r is embedded into the space of continuous functions $C(\mathbb{T})$. In a particular case of W_1^1 we also have embedding into $C(\mathbb{T})$.

In the multivariate case for $\mathbf{x} = (x_1, \dots, x_d)$ denote

$$F_r(\mathbf{x}) := \prod_{j=1}^d F_r(x_j)$$

and

$$\mathbf{W}_p^r := \{f : f = \varphi * F_r, \quad \|\varphi\|_p \leq 1\}.$$

For $f \in \mathbf{W}_p^r$ we will denote $f^{(r)} := \varphi$ where φ is such that $f = \varphi * F_r$.

The main results of Section 2 of [43] are the following two theorems. We use the notation $\beta := \beta(q, p) := 1/q - 1/p$ and $\eta := \eta(q) := 1/q - 1/2$. In the case of trigonometric system \mathcal{T}^d we drop it from the notation:

$$\sigma_m(\mathbf{W})_p := \sigma_m(\mathbf{W}, \mathcal{T}^d)_p.$$

Theorem 5.1. *We have*

$$\sigma_m(\mathbf{W}_q^r)_p \asymp \begin{cases} m^{-r+\beta}(\log m)^{(d-1)(r-2\beta)}, & 1 < q \leq p \leq 2, \quad r > 2\beta, \\ m^{-r+\eta}(\log m)^{(d-1)(r-2\eta)}, & 1 < q \leq 2 \leq p < \infty, \quad r > 1/q, \\ m^{-r}(\log m)^{r(d-1)}, & 2 \leq q \leq p < \infty, \quad r > 1/2. \end{cases}$$

Theorem 5.2. *We have*

$$\sigma_m(\mathbf{W}_q^r)_\infty \ll \begin{cases} m^{-r+\eta}(\log m)^{(d-1)(r-2\eta)+1/2}, & 1 < q \leq 2, \quad r > 1/q, \\ m^{-r}(\log m)^{r(d-1)+1/2}, & 2 \leq q < \infty, \quad r > 1/2. \end{cases}$$

The case $1 < q \leq p \leq 2$ in Theorem 5.1, which corresponds to the first line, was proved in [31] (see also [30], Ch.4). The proofs from [31] and [30] are constructive. In [43] we concentrate on the case $p \geq 2$. We use recently developed techniques on greedy approximation in Banach spaces to prove Theorems 5.1 and 5.2. It is important that greedy approximation allows us not only to prove the above theorems but also to provide a constructive way for building the corresponding m -term approximants. We give a precise formulation from [43].

Theorem 5.3. *For $p \in (1, \infty)$ and $\mu > 0$ there exist constructive methods $A_m(f, p, \mu)$, which provide for $f \in \mathbf{W}_q^r$ an m -term approximation such that*

$$\begin{aligned} & \|f - A_m(f, p, \mu)\|_p \\ & \ll \begin{cases} m^{-r+\beta}(\log m)^{(d-1)(r-2\beta)}, & 1 < q \leq p \leq 2, \quad r > 2\beta + \mu, \\ m^{-r+\eta}(\log m)^{(d-1)(r-2\eta)}, & 1 < q \leq 2 \leq p < \infty, \quad r > 1/q + \mu, \\ m^{-r}(\log m)^{r(d-1)}, & 2 \leq q \leq p < \infty, \quad r > 1/2 + \mu. \end{cases} \end{aligned}$$

Similar modification of Theorem 5.2 holds for $p = \infty$. We do not have matching lower bounds for the upper bounds in Theorem 5.2 in the case of approximation in the uniform norm L_∞ .

As a direct corollary of Theorems 1.1 and 5.1 we obtain the following result.

Theorem 5.4. *Let $p \in [2, \infty)$. Apply the WCGA with weakness parameter $t \in (0, 1]$ to $f \in L_p$ with respect to the real trigonometric system \mathcal{RT}_p^d . If $f \in \mathbf{W}_q^r$, then we have*

$$\|f_m\|_p \ll \begin{cases} m^{-r+\eta}(\log m)^{(d-1)(r-2\eta)+r-\eta}, & 1 < q \leq 2, \quad r > 1/q, \\ m^{-r}(\log m)^{rd}, & 2 \leq q < \infty, \quad r > 1/2. \end{cases}$$

The reader can find results on best m -term approximation as well as results on constructive m -term approximation for Besov type classes in the paper [43]. Some new results in the case of small smoothness are contained in [44].

6 Tensor product approximations

In the paper [1] we study multilinear approximation (nonlinear tensor product approximation) of functions. For a function $f(x_1, \dots, x_d)$ denote

$$\Theta_M(f)_X := \inf_{\{u_j^i\}, j=1, \dots, M, i=1, \dots, d} \|f(x_1, \dots, x_d) - \sum_{j=1}^M \prod_{i=1}^d u_j^i(x_i)\|_X$$

and for a function class F define

$$\Theta_M(F)_X := \sup_{f \in F} \Theta_M(f)_X.$$

In this section we use the notation M instead of m for the number of terms in an approximant because this notation is a standard one in the area. In the case $X = L_p$ we write p instead of L_p in the notation. In other words we are interested in studying M -term approximations of functions with respect to the dictionary

$$\Pi^d := \{g(x_1, \dots, x_d) : g(x_1, \dots, x_d) = \prod_{i=1}^d u^i(x_i)\}$$

where $u^i(x_i)$ are arbitrary univariate functions. We discuss the case of 2π -periodic functions of d variables and approximate them in the L_p spaces. Denote by Π_p^d the normalized in L_p dictionary Π^d of 2π -periodic functions. We say that a dictionary \mathcal{D} has a tensor product structure if all its elements have a form of products $u^1(x_1) \cdots u^d(x_d)$ of univariate functions $u^i(x_i)$, $i = 1, \dots, d$. Then any dictionary with tensor product structure is a subset of Π^d . The classical example of a dictionary with tensor product structure is the d -variate trigonometric system $\{e^{i(\mathbf{k}, \mathbf{x})}\}$. Other examples include the hyperbolic wavelets and the hyperbolic wavelet type system \mathcal{U}^d defined in [35].

Modern problems in approximation, driven by applications in biology, medicine, and engineering, are being formulated in very high dimensions,

which brings to the fore new phenomena. For instance, partial differential equations in a phase space of large spacial dimensions (e.g. Schrödinger and Fokker-Plank equations) are very important in applications. It is known (see, for instance, [4]) that such equations involving large number of spacial variables pose a serious computational challenge because of the so-called *curse of dimensionality*, which is caused by the use of classical notions of smoothness as the regularity characteristics of the solution. The authors of [4] show that replacing the classical smoothness assumptions by structural assumptions in terms of sparsity with respect to the dictionary Π^d , they overcome the above computational challenge. They prove that the solutions of certain high-dimensional equations inherit sparsity, based on tensor product decompositions, from given data. Thus, our algorithms, which provide good sparse approximation with respect to Π^d for individual functions might be useful in applications to PDEs of the above type. The nonlinear tensor product approximation is very important in numerical applications. We refer the reader to the monograph [13] which presents the state of the art on the topic. Also, the reader can find a very recent discussion of related results in [27].

In the case $d = 2$ the multilinear approximation problem is the classical problem of bilinear approximation. In the case of approximation in the L_2 space the bilinear approximation problem is closely related to the problem of singular value decomposition (also called Schmidt expansion) of the corresponding integral operator with the kernel $f(x_1, x_2)$. There are known results on the rate of decay of errors of best bilinear approximation in L_p under different smoothness assumptions on f . We only mention some known results for classes of functions with mixed smoothness. We study the classes \mathbf{W}_q^r of functions with bounded mixed derivative defined above in Section 5.

The problem of estimating $\Theta_M(f)_2$ in case $d = 2$ (best M -term bilinear approximation in L_2) is a classical one and was considered for the first time by E. Schmidt [26] in 1907. For many function classes F an asymptotic behavior of $\Theta_M(F)_p$ is known. For instance, the relation

$$\Theta_M(\mathbf{W}_q^r)_p \asymp M^{-2r+(1/q-\max(1/2,1/p))_+} \quad (6.1)$$

for $r > 1$ and $1 \leq q \leq p \leq \infty$ follows from more general results in [32]. In the case $d > 2$ almost nothing is known. There is (see [33]) an upper estimate in the case $q = p = 2$

$$\Theta_M(\mathbf{W}_2^r)_2 \ll M^{-rd/(d-1)}. \quad (6.2)$$

Results of [1] are around the bound (6.2). First of all we discuss the lower

bound matching the upper bound (6.2). In the case $d = 2$ the lower bound

$$\Theta_M(W_p^r)_p \gg M^{-2r}, \quad 1 \leq p \leq \infty, \quad (6.3)$$

follows from more general results in [32] (see (6.1) above). A stronger result

$$\Theta_M(W_\infty^r)_1 \gg M^{-2r} \quad (6.4)$$

follows from Theorem 1.1 in [34].

We could not prove the lower bound matching the upper bound (6.2) for $d > 2$. Instead, we proved in [1] a weaker lower bound. For a function $f(x_1, \dots, x_d)$ denote

$$\Theta_M^b(f)_X := \inf_{\{u_j^i\}, \|u_j^i\|_X \leq b \|f\|_X^{1/d}} \|f(x_1, \dots, x_d) - \sum_{j=1}^M \prod_{i=1}^d u_j^i(x_i)\|_X$$

and for a function class F define

$$\Theta_M^b(F)_X := \sup_{f \in F} \Theta_M^b(f)_X.$$

In [1] we proved the following lower bound (see Corollary 2.2)

$$\Theta_M^b(\mathbf{W}_\infty^r)_1 \gg (M \ln M)^{-\frac{rd}{d-1}}.$$

This lower bound indicates that probably the exponent $\frac{rd}{d-1}$ is the right one in the power decay of the $\Theta_M(\mathbf{W}_p^r)_p$.

Secondly, we discuss some upper bounds which extend the bound (6.2). The relation (6.1) shows that for $2 \leq p \leq \infty$ in the case $d = 2$ one has

$$\Theta_M(\mathbf{W}_2^r)_p \ll M^{-2r}. \quad (6.5)$$

In [1] we extend (6.5) for $d > 2$.

Theorem 6.1. *Let $2 \leq p < \infty$ and $r > (d-1)/d$. Then*

$$\Theta_M(\mathbf{W}_2^r)_p \ll \left(\frac{M}{(\log M)^{d-1}} \right)^{-\frac{rd}{d-1}}.$$

The proof of Theorem 6.1 in [1] is not constructive. It goes by induction and uses a nonconstructive bound in the case $d = 2$, which is obtained in [33]. The corresponding proof from [33] uses the bounds for the Kolmogorov width $d_n(W_2^r, L_\infty)$, proved by Kashin [15]. Kashin's proof is a probabilistic one, which provides existence of a good linear subspace for approximation, but there is no known explicit constructions of such subspaces. This problem is related to a problem from compressed sensing on construction of good matrices with Restricted Isometry Property. It is an outstanding difficult open problem. In [1] we discuss constructive ways of building good multilinear approximations. The simplest way would be to use known results about M -term approximation with respect to special systems with tensor product structure. However, this approach (see [35]) provides error bounds, which are not as good as best m -term approximation with respect to Π^d (we have exponent r instead of $\frac{rd}{d-1}$ for Π^d). It would be very interesting to provide a constructive multilinear approximation method with the same order of the error as the best m -term approximation.

As we pointed out in a discussion of Theorem 6.1 the upper bound in Theorem 6.1 is proved with a help of probabilistic results. There is no known deterministic constructive methods (theoretical algorithms), which provide the corresponding upper bounds. In [1] we apply greedy type algorithms to obtain upper estimates of $\Theta_M(\mathbf{W}_2^r)_p$. The important feature of our proof is that it is deterministic and moreover it is constructive. Formally, the optimization problem

$$\Theta_M(f)_X := \inf_{\{u_j^i\}, j=1, \dots, M, i=1, \dots, d} \|f(x_1, \dots, x_d) - \sum_{j=1}^M \prod_{i=1}^d u_j^i(x_i)\|_X$$

is deterministic: one needs to minimize over u_j^i . However, minimization by itself does not provide any upper estimate. It is known (see [5]) that simultaneous optimization over many parameters is a very difficult problem. Thus, in nonlinear M -term approximation we look for methods (algorithms), which provide approximation close to best M -term approximation and at each step solve an optimization problem over only one parameter ($\prod_{i=1}^d u_j^i(x_i)$ in our case). In [1] we provide such an algorithm for estimating $\Theta_M(f)_p$. We call this algorithm *constructive* because it provides an explicit construction with feasible one parameter optimization steps. We stress that in the setting of approximation in an infinite dimensional Banach space, which is considered in [1], the use of term *algorithm* requires some explanation. In that paper

we discuss only theoretical aspects of the efficiency (accuracy) of M -term approximation and possible ways to realize this efficiency. The *greedy algorithms* used in [1] give a procedure to construct an approximant, which turns out to be a good approximant. The procedure of constructing a greedy approximant is not a numerical algorithm ready for computational implementation. Therefore, it would be more precise to call this procedure a *theoretical greedy algorithm* or *stepwise optimizing process*. Keeping this remark in mind we, however, use the term *greedy algorithm* in this paper because it has been used in previous papers and has become a standard name for procedures used in [1] and for more general procedures of this type (see for instance [6], [40]). Also, the theoretical algorithms, which we use in [1], become algorithms in a strict sense if instead of an infinite dimensional setting we consider a finite dimensional setting, replacing, for instance, the L_p space by its restriction on the set of trigonometric polynomials. We note that the greedy-type algorithms are known to be very efficient in numerical applications (see, for instance, [47] and [25]).

In [1] we use two very different greedy-type algorithms to provide a constructive multilinear approximant. The first greedy-type algorithm is based on a very simple dictionary consisting of shifts of the de la Vallée Poussin kernels. The algorithm uses function (dyadic blocks of a function) evaluations and picks the largest of them. The second greedy-type algorithm is more complex. It is based on the dictionary Π^d and uses the Weak Chebyshev Greedy Algorithm with respect to Π^d to update the approximant. Surprisingly, these two algorithms give the same error bound. For instance, Theorems 4.3 and 4.4 from [1] give for big enough r the following constructive upper bound for $2 \leq p < \infty$

$$\Theta_M(\mathbf{W}_2^r)_p \ll \left(\frac{M}{(\ln M)^{d-1}} \right)^{-\frac{rd}{d-1} + \frac{\beta}{d-1}}, \quad \beta := \frac{1}{2} - \frac{1}{p}.$$

This constructive upper bound has an extra term $\frac{\beta}{d-1}$ in the exponent compared to the best M -term approximation. It would be interesting to find a constructive proof of Theorem 6.1.

References

- [1] D. Bazarkhanov and V. Temlyakov, Nonlinear tensor product approximation of functions, arXiv:1409.1403v1 [stat.ML] 4 Sep 2014 (to appear in J. Complexity, 2015).
- [2] E.S. Belinskii, Decomposition theorems and approximation by a "floating" system of exponentials, Transactions of the American Mathematical Society, **350** (1998), 43–53.
- [3] W. Dai and O. Milenkovic, Subspace pursuit for compressive sensing signal reconstruction, IEEE Trans. Inf. Theory, **55** (2009), 2230–2249.
- [4] W. Dahmen, R. DeVore, L. Grasedyck, and E. Süli, Tensor-sparsity of solutions to high-dimensional elliptic Partial Differential Equations, Manuscript, July 23, 2014.
- [5] G. Davis, S. Mallat, and M. Avellaneda, Adaptive greedy approximations, Constructive Approximation **13** (1997), 57–98.
- [6] R.A. DeVore, Nonlinear approximation, Acta Numerica, **7** (1998), 51–150.
- [7] R.A. DeVore and V.N. Temlyakov, Nonlinear approximation by trigonometric sums, J. Fourier Analysis and Applications, **2** (1995), 29–48.
- [8] S.J. Dilworth, N.J. Kalton, and Denka Kutzarova, On the existence of almost greedy bases in Banach spaces, Studia Math., **158** (2003), 67–101.
- [9] S.J. Dilworth, D. Kutzarova, V.N. Temlyakov, Convergence of some Greedy Algorithms in Banach spaces, *The J. Fourier Analysis and Applications* **8** (2002), 489–505.
- [10] S.J. Dilworth, M. Soto-Bajo and V.N. Temlyakov, Quasi-greedy bases and Lebesgue-type inequalities, Stud. Math., **211** (2012), 41–69.
- [11] G. Garrigós, E. Hernández, and T. Oikhberg, Lebesgue type inequalities for quasi-greedy bases, Constr. Approx., **38** (2013), 447–479.

- [12] E.D. Gluskin, Extremal properties of orthogonal parallelepipeds and their application to the geometry of Banach spaces, *Math USSR Sbornik* **64** (1989), 85–96.
- [13] W. Hackbusch, Tensor Spaces and Numerical Tensor Calculus, Springer-Verlag, Heidelberg, 2012.
- [14] R.S. Ismagilov, Widths of sets in normed linear spaces and the approximation of functions by trigonometric polynomials, *Uspekhi Mat. Nauk*, **29** (1974), 161–178; English transl. in *Russian Math. Surveys*, **29** (1974).
- [15] B.S. Kashin, Widths of certain finite-dimensional sets and classes of smooth functions, *Izv. AN SSSR*, **41** (1977), 334–351; English transl. in *Math. Izv.* **11** (1977).
- [16] B.S. Kashin and V.N. Temlyakov, On best m -term approximations and the entropy of sets in the space L^1 , *Math. Notes*, **56** (1994), 1137–1157.
- [17] S.V. Konyagin and V.N. Temlyakov, A remark on greedy approximation in Banach spaces, *East. J. Approx.* **5** (1999), 365–379.
- [18] E.D. Livshitz and V.N. Temlyakov, Sparse approximation and recovery by greedy algorithms, *IEEE Transactions on Information Theory*, **60** (2014), 3989–4000; arXiv:1303.3595v1 [math.NA] 14 Mar 2013.
- [19] V.E. Maiorov, Trigonometric diameters of the Sobolev classes W_p^r in the space L_q , *Math. Notes* **40** (1986), 590–597.
- [20] Y. Makovoz, On trigonometric n -widths and their generalizations, *J. Approx. Theory* **41** (1984), 361–366.
- [21] D. Needell and J.A. Tropp, CoSaMP: iterative signal recovery from incomplete and inaccurate samples, *Applied and Computational Harmonic Analysis*, **26** (2009), 301–321.
- [22] D. Needell and R. Vershynin, Uniform uncertainty principle and signal recovery via orthogonal matching pursuit, *Found. Comp. Math.*, **9** (2009), 317–334.

- [23] M. Nielsen, An example of an almost greedy uniformly bounded orthonormal basis for $L_p(0, 1)$, *J. Approx. Theory*, **149** (2007), 188–192.
- [24] A.S. Romanyuk, Best M -term trigonometric approximations of Besov classes of periodic functions of several variables, *Izvestia RAN, Ser. Mat.* **67** (2003), 61–100; English transl. in *Izvestiya: Mathematics* (2003), 67(2):265.
- [25] S. Shalev-Shwartz, N. Srebro, and T. Zhang, Trading accuracy for sparsity in optimization problems with sparsity constraints, *SIAM Journal on Optimization*, **20(6)** (2010), 2807–2832.
- [26] E. Schmidt, Zur Theorie der linearen und nichtlinearen Integralgleichungen. I. *Math. Ann.*, **63** (1907), 433–476.
- [27] R. Schneider and A. Uschmajew, Approximation rates for the hierarchical tensor format in periodic Sobolev spaces, *J. Complexity* **30** (2014), 56–71.
- [28] S. B. Stechkin, On absolute convergence of orthogonal series, *Dokl. AN SSSR*, **102** (1955), 37–40 (in Russian).
- [29] D. Savu and V.N. Temlyakov, Lebesgue-Type Inequalities for Greedy Approximation in Banach Spaces, *IEEE Transactions on Information Theory*, **58** (2013), 1098–1106.
- [30] V.N. Temlyakov, Approximation of functions with bounded mixed derivative, *Trudy MIAN*, **178** (1986), 1–112. English transl. in *Proc. Steklov Inst. Math.*, **1** (1989).
- [31] V.N. Temlyakov, Approximation of Periodic Functions of Several Variables by Bilinear Forms, *Izvestiya AN SSSR*, **50** (1986), 137–155; English transl. in *Math. USSR Izvestija*, **28** (1987), 133–150.
- [32] V.N. Temlyakov, Estimates of the best bilinear approximations of functions of two variables and some of their applications, *Mat. Sb.*, **134** (1987), 93–107; English transl. in *Math. USSR-Sb* **62** (1989), 95–109.
- [33] V.N. Temlyakov, Estimates of best bilinear approximations of periodic functions, *Trudy Mat. Inst. Steklov*, **181** (1988), 250–267; English transl. in *Proc. Steklov Inst. of Math.* **4** (1989), 275–293.

- [34] V.N. Temlyakov, Estimates of Best Bilinear Approximations of Functions and Approximation Numbers of Integral Operators, *Matem. Zametki*, **51** (1992), 125–134; English translation in *Math. Notes*, **51** (1992), 510–517.
- [35] V.N. Temlyakov, Greedy Algorithms with Regard to Multivariate Systems with Special Structure, *Constr. Approx.*, **16** (2000), 399–425
- [36] V.N. Temlyakov, Nonlinear Kolmogorov’s widths, *Matem. Zametki*, **63** (1998), 891–902.
- [37] V.N. Temlyakov, Greedy algorithms in Banach spaces, *Adv. Comput. Math.*, **14** (2001), 277–292.
- [38] V.N. Temlyakov, Nonlinear method of approximation, *Found. Comput. Math.*, **3** (2003), 33–107.
- [39] V.N. Temlyakov, Greedy-Type Approximation in Banach Spaces and Applications, *Constr. Approx.* **21** (2005), 257–292.
- [40] V.N. Temlyakov, *Greedy approximation*, Cambridge University Press, 2011.
- [41] V.N. Temlyakov, Sparse approximation and recovery by greedy algorithms in Banach spaces, *Forum of Mathematics, Sigma*, **2** (2014), e12, 26 pages; arXiv:1303.6811v1 [stat.ML] 27 Mar 2013, 1–27.
- [42] V.N. Temlyakov, An inequality for the entropy numbers and its application, *J. Approx. Theory*, **173** (2013), 110–121.
- [43] V.N. Temlyakov, Constructive sparse trigonometric approximation and other problems for functions with mixed smoothness, arXiv: 1412.8647v1 [math.NA] 24 Dec 2014, 1–37 (to appear in *Math. Sbornik*, 2015).
- [44] V.N. Temlyakov, Constructive sparse trigonometric approximation for functions with small mixed smoothness, arXiv: 1503.0282v1 [math.NA] 1 Mar 2015, 1–30.
- [45] V.N. Temlyakov, Mingrui Yang and Peixin Ye, Greedy approximation with regard to non-greedy bases, *Adv. Comput. Math.* **34** (2011), 319–337.

- [46] P. Wojtaszczyk, Greedy Algorithm for General Biorthogonal Systems, J. Approx. Theory **107** (2000), 293-314.
- [47] T. Zhang, Sequential greedy approximation for certain convex optimization problems, IEEE Transactions on Information Theory, **49(3)** (2003), 682–691.
- [48] T. Zhang, Sparse Recovery with Orthogonal Matching Pursuit under RIP, IEEE Transactions on Information Theory, **57** (2011), 6215–6221.